Library and Information Services in Astronomy VII
ASP Conference Series, Vol. 492
Andras Holl, Soizick Lesteven, Dianne Dietrich, and Antonella Gasperini, eds.

Bibliographic Data in Astronomy: Experience with the IBVS Reference List Revision

A. Holl

Konkoly Observatory, H-1525 P.O. Box 67, Budapest, Hungary &
Library and Information Centre, Hungarian Academy of Sciences

Abstract. Literature has to be both visible and accessible. What is not on the web is almost non-existent. Let's turn this adage around: are our present bibliographies complete? Where are the blind spots? Are there resources missing from the web, or not readily accessible?

The author shares his experience gained during the extensive revision of old reference lists from the Information Bulletin on Variable Stars (IBVS). ADS contains about twenty-five thousand references from IBVS issues between 1961 and 2013. There are some more references in journals, unidentifiable by ADS. Some are incomplete or inaccurate, and the rest is mostly old and obscure. But however old or obscure it is, it must contain important information, because it is cited. Old observatory publications, and aged gray literature in general, is just in the process of being cleared off from library shelves. It is not only the literature of the past we need to discuss — there are challenges for the present and the future: these include new forms of publications that are hard to render into bibcodes, data and data products, and items that are not strictly data or literature, like VOEvents and nanopublications.

Introduction

The Information Bulletin on Variable Stars (IBVS) is a small, Open Access electronic journal specializing in variable star research, published by Konkoly Observatory, Budapest. IBVS has many enhanced features (Holl 2012). One of those is reference linking. Using a technology based on software developed at the Centre de Données astronomiques (CDS), Strasbourg, reference links are created on the fly from the LaTeX source of the article. It is made possible by the bibcode system used in astronomy (Schmitz et al. 1995). Bibcodes are computable, which means that they can be constructed from the references in an article, most of the time. In contrast, DOI linking used by CrossRef members involves insertion of links *a posteriori*, as DOI suffixes are not necessarily computable.

Bibcode creation works for "born digital" issues of IBVS that have been published since 1994. And it works very well, partly because the target bibliographic database — the ADS — has an almost complete coverage of astronomical literature, including select textbooks on Mathematics and Physics. Therefore, not many publications cited in IBVS are missing from the database. The other reason for this high level of efficiency is the linking itself: because references could be checked with little manual labor, those that could not be automatically linked will be tagged with the appropriate bibcode manually. This system has significantly improved the accuracy of references in the journal.

Issues with IBVS reference linking

IBVS reference linking requires the use of a consistent, yet flexible citation style. While the order of semantic citation elements is fixed — there are rules for delimiters, for instance — using a fixed vocabulary of journal names is not necessary. Presently, about 150 different periodicals are recognized in about 600 name variants or abbreviations. Conference proceedings are often cited by using the same style as journals, using a series name and number: this allows for the consistent creation of bibcodes for these types of materials. Bibcodes are sometimes inserted in the LaTeX source by the authors, though most of the time it is unnecessary. We create links from URLs, VizieR dataset identifiers, arXiv preprint IDs, and DOIs as well.

The HTML version of the journal has good reasons to use automatic link generation. IBVS is a peer reviewed express journal, covering a field where rapid turnaround times are necessary. After a quick editorial check, the process of editing reference links does not delay the process. Articles appear online rapidly, and the links are there instantly, requiring little manpower.

During 2012–13, we revised the IBVS issues published before 1994, which are available online as OCR-ed text files. Previously only local links — references to IBVS or Konkoly Observatory Communications articles — were available for those old issues, and other publications were not linked. The non-consistent manuscript style and the unreliable references in old issues made reference linking prohibitive up till now. The references were unreliable, because the editors did not have the tools that we have now, such as the reference link generator and the ADS. Reference syntax varied and was intended for humans, not machines.

To further complicate the issue, astronomers often cited the page number for a table instead of the first page of the article, and the observer instead of the author of the paper. (We have to note that ADS bibcodes for old literature are less tidy too.) Finally, the text, despite several earlier revisions, contains OCR errors, some of which are in references. Not discouraged by those difficulties, we have started a revision, and manually inserted bibcodes to the text, from which links are created in the web edition of the journal.

Using ADS to identify references

During this project, we have relied on ADS exclusively; the library of Konkoly Observatory was not used at all. All references we could not identify in ADS — for any reason — were skipped. Before the revision, ADS contained about 22000 references from IBVS. After, there was a 10% increase in items that were previously unknown by the database. There were slightly more than 10000 bibcodes inserted into 4100 IBVS issues.

During the process, a considerable number of errors were detected both in IBVS (which were OCR errors) and ADS. Errors in IBVS were corrected and errors in ADS were reported. We were surprised that some IBVS issues were missing from ADS: these were supplied, obviously. (Issues of IAU Circulars were found to be missing from ADS as well.) There were expected gaps in ADS — Russian variable star literature from the 1930s or old observatory publications, especially non-English, are often missing. The number of unidentifiable references in ADS increases with age.

ADS uses a different mechanism for processing the references. The results, unsurprisingly, differ too: the list of references for a given IBVS issue might be different in the ADS database from the list of links in the HTML version on the web. For the scanned IBVS issues from the period before 1994, any false references were reported to ADS as a result of the revision project. Moving forward, we will try to pass verified bibcodes to ADS for current issues: there is no need to resolve references that were already resolved and verified at the journal's side.

At this point, the reader might get the impression that reference linking for the current issues of IBVS has no challenges. This is, unfortunately, not quite true. Changes in scientific publishing — astronomy publishing in particular — might render this elegant method of reference linking less effective. The number of incomputable bibcodes is increasing as the bibcode usage starts to deviate from the old standard. There might be good reasons for this change, but it might endanger on-the-fly bibcode generation nevertheless.

Issues with referencing additional types of research outputs

Let's take a few steps back, put IBVS into perspective, and take a look at reference linking and bibliographic hubs in general. Most astronomy journals have reference links in the HTML versions of papers. Most scholarly journals in science use the CrossRef system for reference linking, with the DOI database acting as a hub. Most publishers use an internal, "in house" linking system and most astronomy journals link to ADS. Few journal articles remain unlinked: those references with no links are mostly for conference proceedings and books. Articles in science refer mostly to other journal articles and to conference papers.

Links are often missing for gray literature. "Private communications" remain unlinked. We have to admit a certain amount of frustration seeing the authors use such references so frequently. Though these might be unavoidable occasionally, we think these should be the exceptions. Why cite something that cannot be accessed? What contains useful (citable) information should be accessible online. With repositories being available, such documentation (including theses) should not present a problem any longer.

"In press" articles, understandably, have no links. At IBVS, we made a feeble effort to supply bibcodes retroactively. We think this practice improves the value of the article. When arXiv preprints are cited and linked, ADS might connect the preprint to the final published article.

Value of older materials

Astronomical bibliographies have come a long way since Weidler's. Today ADS is, undoubtedly, "the" astronomy bibliography. It does not mean that ADS is complete, but we believe it is more complete within the field than its commercial competitors. Where are its blind spots? We say, based on the experience with IBVS, mostly the old, non-English, non-journal (e.g., observatory publications, books, documentation) material is missing.

Do we have any chance to fill in those gaps? We see a danger of losing this part of the astronomy literature, as astronomy libraries are being re-organized and old obser-

vatory publications sometimes moved to storage or being discarded. Those materials not in ADS should be cut up and scanned, and discarded only afterward digitization.

Does old and obscure equal useless? Apparently not, because such articles are being still being cited. Journal citing half-life does vary from field to field, discipline to discipline. In variable star research, where long time variability might be a subject of research, old articles are cited frequently. IBVS reference and citation half-lives are about seven years. The Figure 1 shows the age distribution of citations in a sample of IBVS issues spreading to four decades, and the age distribution of citations the same IBVS issues received. There are publications older than 50 years cited in IBVS articles and some articles published in IBVS are still being cited after 50 years.

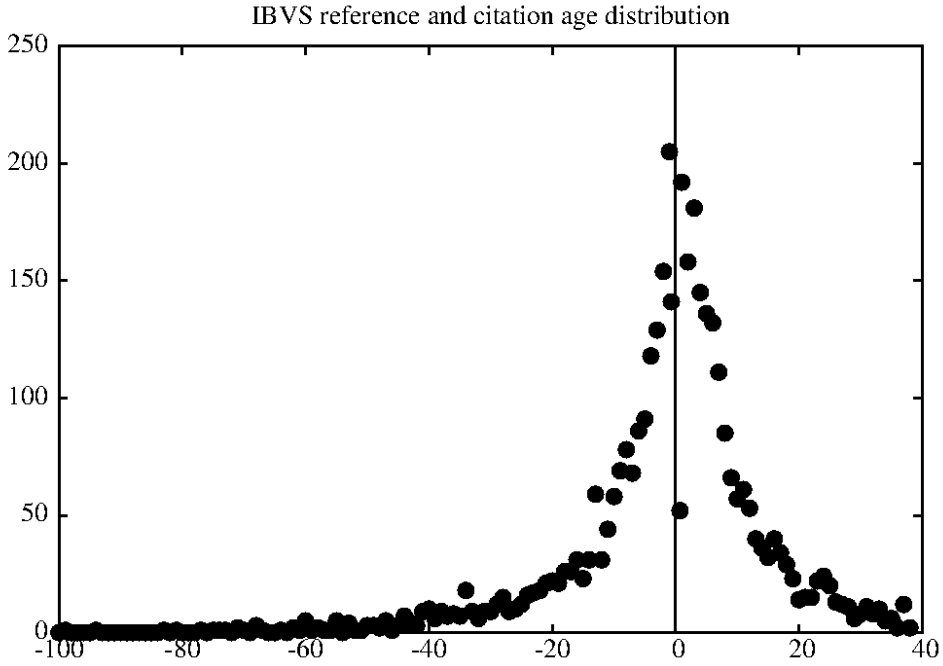

Figure 1. Age distribution of references in, and citations to IBVS. The labels shown below the horizontal axis represent years, and the vertical bar at year zero marks the time of publication. Reference age distribution is shown on the left side, and citation age distribution on the right side. IBVS was founded in 1961, hence the upper limit of citation ages.

Books represent only a minority of referred sources, and their data is often missing from ADS. In certain areas, open access book publishing is on the rise. We hope astronomy books will be more often available online and will be represented in ADS. Non-English literature, and literature in languages using a non-Latin script are cited rarely, if at all now in IBVS. Sources in Russian were often cited several decades ago. Based on the citation evidence of IBVS, Russian astronomical literature should be processed and bibliographical data fed to ADS.

Implications for emerging research output types

The IBVS method of generating bibcodes from references is rather a peculiarity than the norm. Nevertheless, we feel that it would be wrong to violate the old bibcode rules and squeeze those publications with too long identifiers into the old pattern. To provide one example for illustration: *EPJ Web of Conferences* uses 5 digit identifiers, which refuse to fit into the slot of either page numbers or volumes. DOIs allow the use of long numbers, while bibcodes do not. Maybe the bibcode standard should be changed instead of hacked.

More of this might come along. We wonder whether VOEvents or nanopublications will be cited in journal articles. Methods for referring to datasets are evolving, and we could mention the Universal Numeric Fingerprint (Altman and King 2007). ADS contains datasets and uses bibcodes for those. We will see how ADS will cope with such novel types of publication.

Conclusion

For an overview of astronomical bibliographies, see Duerbeck (2005). The first bibliography he mentions, that of Johann Friedrich Weidler (Weidler 1755), was published in a re-print edition in 2009, for the benefit of collectors of old astronomical books. The Astronomischer Jahresbericht and the triple-A together covered a century: some of that material is available through ARIBIB. We think that old bibliographies should be made available in online databases. New forms of publications will undoubtedly be covered by bibliographies. Astronomers need those services — the ocean of literature would be unnavigable without them.

References

Altman, M., King, G. 2007, D-Lib Magazine, 13, 3/4, http://www.dlib.org/dlib/march07/altman/03altman.html
Duerbeck, H.W. 2005, ASSL, 330, 153
Holl, A. 2012, D-Lib Magazine, 18, 5/6, http://www.dlib.org/dlib/may12/holl/05holl.html
Schmitz, M., Helou, G., Lague, C., Madore, B., Corwin, G. G., Jr., Dubois, P., & Lesteven, S. 1995, Vistas in Astronomy, 39, 272
Weidler, J.F. 1755, Bibliographia Astronomica Temporis